Response Under 37 CFR §1.111
PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.	:	10/044,720	Confirmation No. 1251
Applicant	:	AMIR et al.	
Filed	:	01/11/2002	
TC/A.U.	:	2171	
Examiner	:	Cam Linh T. NGUYEN	
Docket No.	:	ARC920000131US1	
Customer No.	:	23334	

37 C.F.R. 1.131 DECLARATION

I, the undersigned, inventor of the above-referenced patent application, hereby declare the following:

1) The pending claims of our above identified patent invention were rejected under 35 U.S.C. §102(e) / § 103 based on the prior art references of Ingle et al. (U.S. Patent Pub. No. 2002/0138524) with a filing date of January 19, 2001. (hereinafter referred to as "**Ingle**") In view of Change et al. (U.S. Patent Pub. No. 2003/0050923) with a filing date of December 21, 2001 (hereinafter referred to as "**Chang**").
2) The invention described in the above referenced patent application was reduced to a writing and signed by the undersigned applicants prior to the December 21, 2001 filing date of **Chang**, and further prior to the January 19, 2001 filing date of **Ingle**. In particular, the relevant portion of our Invention Disclosure upon which the above referenced patent application was based is attached herewith.

I, the undersigned, declare all of the above statements are made on my own knowledge, the above statements are true and correct, and the above statements are made on information that I believe to be true. I understand that false statements or concealment in obtaining a patent will subject me to fine and/or imprisonment or both (18 U.S.C. §1001) and may jeopardize the validity of the above identified patent application or any application issuing therefrom.

Amon AMIR	Gal ASHOUR	Matthew DENESUK
November 3rd 2004	November ___, 2004	November ___, 2004

NOTE: Inventors Brian K. BLANCHARD and Reiner KRAFT are unavailable for signature under MPEP § 715.04

Response Under 37 CFR §1.111
PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.	:	10/044,720	Confirmation No. 1251
Applicant	:	AMIR et al.	
Filed	:	01/11/2002	
TC/A.U.	:	2171	
Examiner	:	Cam Linh T. NGUYEN	
Docket No.	:	ARC920000131US1	
Customer No.	:	23334	

37 C.F.R. 1.131 DECLARATION

I, the undersigned, inventor of the above-referenced patent application, hereby declare the following:

1) The pending claims of our above identified patent invention were rejected under 35 U.S.C. §102(e) / § 103 based on the prior art references of Ingle et al. (U.S. Patent Pub. No. 2002/0138524) with a filing date of January 19, 2001. (hereinafter referred to as "**Ingle**") in view of Change et al. (U.S. Patent Pub. No. 2003/0050923) with a filing date of December 21, 2001 (hereinafter referred to as "**Chang**").
2) The invention described in the above referenced patent application was reduced to a writing and signed by the undersigned applicants prior to the December 21, 2001 filing date of **Chang**, and further prior to the January 19, 2001 filing date of **Ingle**. In particular, the relevant portion of our Invention Disclosure upon which the above referenced patent application was based is attached herewith.

I, the undersigned, declare all of the above statements are made on my own knowledge, the above statements are true and correct, and the above statements are made on information that I believe to be true. I understand that false statements or concealment in obtaining a patent will subject me to fine and/or imprisonment or both (18 U.S.C. §1001) and may jeopardize the validity of the above identified patent application or any application issuing therefrom.

Amon AMIR	Gal ASHOUR	Matthew DENESUK
November __, 2004	November 3, 2004	November ___, 2004

NOTE: Inventors Brian K. BLANCHARD and Reiner KRAFT are unavailable for signature under MPEP § 715.04

Response Under 37 CFR §1.111
PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.	:	10/044,720	Confirmation No. 1251
Applicant	:	AMIR et al.	
Filed	:	01/11/2002	
TC/A.U.	:	2171	
Examiner	:	Cam Linh T. NGUYEN	
Docket No.	:	ARC920000131US1	
Customer No.	:	23334	

37 C.F.R. 1.131 DECLARATION

I, the undersigned, inventor of the above-referenced patent application, hereby declare the following:

1) The pending claims of our above identified patent invention were rejected under 35 U.S.C. §102(e) / § 103 based on the prior art references of Ingle et al. (U.S. Patent Pub. No. 2002/0138524) with a filing date of January 19, 2001. (hereinafter referred to as "**Ingle**") in view of Change et al. (U.S. Patent Pub. No. 2003/0050923) with a filing date of December 21, 2001 (hereinafter referred to as "**Chang**").
2) The invention described in the above referenced patent application was reduced to a writing and signed by the undersigned applicants prior to the December 21, 2001 filing date of **Chang**, and further prior to the January 19, 2001 filing date of **Ingle**. In particular, the relevant portion of our Invention Disclosure upon which the above referenced patent application was based is attached herewith.

I, the undersigned, declare all of the above statements are made on my own knowledge, the above statements are true and correct, and the above statements are made on information that I believe to be true. I understand that false statements or concealment in obtaining a patent will subject me to fine and/or imprisonment or both (18 U.S.C. §1001) and may jeopardize the validity of the above identified patent application or any application issuing therefrom.

Amon AMIR	Gal ASHOUR	Matthew DENESUK
November ___, 2004	November ___, 2004	November 4, 2004

NOTE: Inventors Brian K. BLANCHARD and Reiner KRAFT are unavailable for signature under MPEP § 715.04

Disclosure

Created By: Gal Ashour **Created On:**
Last Modified By: Matthew Denesuk **Last Modified On:**

*** IBM Confidential ***

Required fields are marked with the asterisk (*) and must be filled in to complete the form.

Summary

Status	
Processing Location	
Functional Area	
Attorney/Patent Professional	
IDT Team	
Submitted Date	
Owning Division	
Select	
	To calculate a PVT score, use the 'Calculate PVT' button.
Incentive Program	
Lab	
Technology Code	

Inventors with Lotus Notes IDs

Inventors: Gal Ashour/Almaden/IBM, Arnon Amir/Almaden/IBM, Matthew Denesuk/Almaden/IBM, Reiner Kraft/Almaden/IBM, Brian Blanchard/Almaden/IBM

Inventor Name > denotes primary contact	Inventor Serial	Div/Dept	Manager Serial	Manager Name
Ashour, Gal				
Amir, Arnon				
> Denesuk, Matthew				
Kraft, Reiner				
Blanchard, Brian K.				

Inventors without Lotus Notes IDs

IDT Selection

IDT Team:	Attorney/Patent Professional:

Response Due to IP&L :

Main Idea

***Title of disclosure (in English)**
Synthesizing information-bearing content from multiple channels

***Idea of disclosure**
1. Describe your invention, stating the problem solved (if appropriate), and indicating the advantages of using the invention.

The present invention addresses the twofold problem of quantity and quality associated with the explosive growth on content available via the internet. People are increasingly finding it difficult to sort through the great mass of content available with apparent relevance to even a narrowly-defined need. They are overwhelmed with information and greatly desire appropriate tools or services that can aggregate relevant content and produce appropriate distillations or summarizations. Further, the ease with which content can be made available on the internet, along with the greater speed with which "raw" content is generally made available, has created a greater-than-ever need for efficient means of judging the correctness of information. Strictly human-based editorial judgement is becoming no longer viable nor desirable.

The present invention leverages the growing capabilities of computational power and available bandwidth to automate both the functions of content aggregation/summarization and editorial judgement. The present invention further leverages technology related to automated video indexing and content extraction (e.g., via speech-to-text or phoneme recognition) to allow for one to manipulate, evaluate, and judge video-based content.

2. How does the invention solve the problem or achieve an advantage,(a description of "the invention", including figures inline as appropriate)?
The present invention provides means for synthesizing information-bearing content from multiple channels. The central concept relates to executing analyses of mutually similar elements of content (EOC's) to generate "digests," or new EOC's which reflect the degree of corroboration (DOC) or other comparative measure between the original EOC's.

The original EOC's and the derived digests may comprise text, video, audio, other media formats, or any combination thereof.

The channels may comprise online news sources, video broadcasts, press release forums, financial forums, etc.

The present invention will enable users to search for, access, and use improved and/or more convenient forms of content, such as summaries, synopses possessing a high degree of corroboration, update briefings, etc.

Datastream sources can be located by suitable crawlers, or they can be identified manually. After identification, it is necessary to "wrap" or create an interface which enables the extraction of individual EOC's from that source. This task is generally being made easier by the adoption of XML standards, but it still frequently requires an amount of manual programming. Once the "wrapper" for that source has been programmed, however, it should reliably extract EOC's until the owner of the source changes its interface.

It is generally preferable to make an arrangement with the owner or controller of the source to obtain a more efficient and robust interface to that source.

To aggregate similar EOC's, create digests/summaries, and execute editorial judgments, it is

necessary to analyze the context and meaning of the content. When EOC's comprise video-based content, it is necessary to create or otherwise obtain or extract metadata which characterizes the video in some meaningful way. Examples of means by which such characterization may be effected include synchronized indexing using the audio track of the video, ocr-based extraction of closed captioning or other textual information contained in the video, or direct extraction from linked or otherwise associated sources (e.g., teletext). If video indexing is used, it is preferred to use highly automated video indexing means such as that associated with IBM's CueVideo Multimedia Project [Give References].

Methods known in the art of Content Management and Knowledge Management are used to cluster EOC's, correlate EOC's, produce summaries, and provide measures of the DOC.

Prior Art

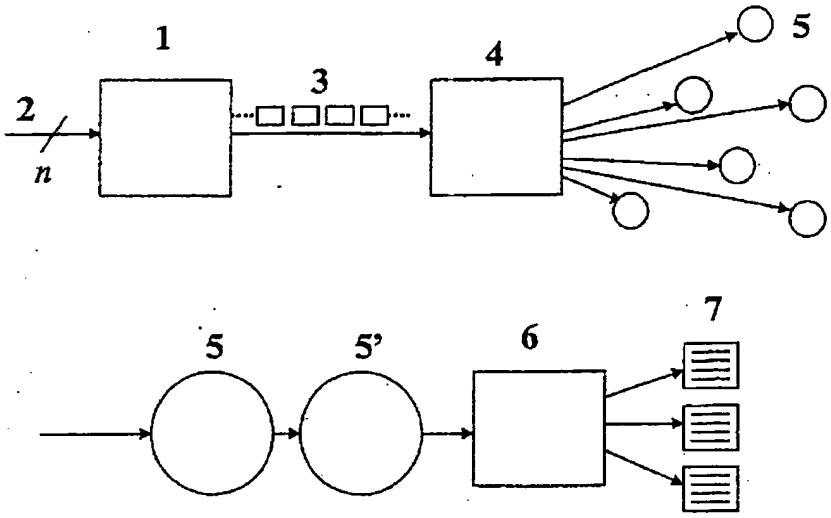
Means for segmenting or extracting discrete elements of content from text-based sources are known in the art and are hereby incorporated by reference. Examples include US#6,052,657, which describes means for segmenting text and identifying topics contained therein that match a user query. US#5,806,021 discloses two statistical methods for segmenting continuous text sources. The first method simple and rapid but is more error-prone than the second. The second, which uses statistical language models, is more reliable and accurate, but it is also more computationally intensive.

Means for summarization of text-based sources are also known in the art and are hereby incorporated by reference. Examples include US#5,918,240, which uses an iterative approach for training a system to select the same summarizing sentences from a document that a human expert would make. The methodology involves feature value probabilities extracted from manually generated summaries performed by human experts. US#5,077,668 describes means of automatically producing abstracts of documents by making use of pre-selected "hint" words, which signal the presence of "significant phrases." US# 4,358,824 performs an abstracting function in which appropriate keywords are automatically extracted using a parts of speech directory. These keywords are then used as the basis for creating a useful abstract. US#5,924,108 describes means for generating a summary from a text-based document based in statistical analysis of the frequency of occurrence of "content" words in sentences or phrases. Higher-ranked sentences or phrases are selected to comprise the summary. In addition, the user may pre-identify words or phrases which he/she specifically wants excluded. Additional programmed sentences may be included in the summary conditioned upon the specific results obtained from the preceding analysis. US#5,689,716 discloses means for automatically generating thematic summaries of documents. These means require the specification or extraction of thematic terms, the corresponding frequencies of occurrence of which are used to rank thematic sentences for inclusion in the thematic summary.

Preferred Embodiment

A preferred embodiment of the invention may be seen as a combination of several filters (see Figure 1 below). The first filter, **1**, has as inputs a plurality of content channels, **2**. The filter performs a function of separating or segmenting the content streams into individual EOC's and "tagging" each with metadata that indicate source, date, and other relevant information. The output of filter **1** is a stream, **3**, of these tagged elements.

Figure 1



A second filter, 4, performs a pattern matching or similar function on the EOC data stream, 3, and outputs EOC's to a set of virtual buffers, 5, where each virtual buffer contains a set of appropriately related EOC's. In a preferred sub-embodiment, a "distance function" is defined so that one can compute a set of "distances" from each EOC to every other EOC. Each of the virtual buffers, 5, is then made to contain an EOC's and every EOC which is less than a certain "distance" from it. Optionally, EOC's may be made to "expire" after a particular threshold (e.g., its age) is exceeded. Naturally, a given EOC may be contained in a plurality of virtual buffers.

The grouping of EOC's by mutual distance will lead to various data structures, depending on the precise nature of the data. For highly heterogeneous data, the grouping may coalesce into sharply-defined and distinct EOC groupings. For less heterogeneous data, however, the boundaries may be fuzzy and it will be non-trivial to define EOC groupings. These problems are commonly addressed in the Prior Art of the fields of Content Management and Knowledge Management, and this Art is herein incorporated by reference.

Optionally, virtual buffers, 5, may be followed by processing to create virtual summary buffers, 5', which serve to create summaries or perform other advantageous preprocessing which supports the comparative analysis performed by the filter, 6.

The content of each virtual buffer, 5, or virtual summary buffer, 5', is passed through a third filter, 6, which performs a comparative analysis of the "mutually close" EOC's, and, based on a set of heuristics, outputs a set of "digests," 7. These digests will comprise EOC formats which benefit from such comparative analyses, such as summaries, highly corroborated "findings of fact," trailers (e.g., the most vibrant or exciting details of a set of EOC's), and updates (e.g., all related information released within the last 48 hours).

An example of a preferred means for producing a set of summary digests, 7, from the virtual buffers, 5 or 5', comprises concatenating the EOC's included in the virtual buffer, and then applying a redundancy filter to remove redundant sub-elements. Sub-elements are taken as being redundant if they are determined to be substantially similar to each other (by knowledge management techniques, e.g., those involving distance functions, known in the art). The acceptable degree of similarity present in order for two sub-elements to be considered similar is determined by setting a threshold, the value of which is used to fine-tune the process of creating the summary digests. Another threshold indicating the degree redundancy which must be present for a representative sub-element to be included in the summary digest may also be included.

A preferred means for presenting the summary digest is to color-code each sub-element of content based on the number of EOC's containing that particular sub-element or sub-elements substantially similar to that particular sub-element. For example, a red font might indicate that content substantially similar to that sub-element appeared in greater than 50% of the relevant EOC's; a yellow font might indicted the corresponding appearance was between 25% and 50%; a blue font might indicate appearance between 5% and 25%; and a gray font might indicate appearance less than 5%.

Users are able to search on these digests, **7**, based on their content, the metadata they carry, and their format.

The digests are updated on a dynamic basis.

In addition to the format of the digest, the user may specify an approximate length of the digest. This will factor into thresholding characteristics of the filter, **6**, which produces the digests from the dynamic content in the virtual buffers.

The digests contain both aggregate and more individualized metadata. For example, the digest as a whole will be assigned a "degree of corroboration" (DOC), which provides a measure of how highly corroborated the content is. Furthermore, users may query with regard to the effective DOC of individual components of the digest, or even be directed to some of the original sources. The metadata structure must be designed to enable such functionality.

A database or index may be maintained which reflects the historical DOC data of content available from individual publishers or other information providers or aggregators. These historical data are used to weight future DOC data appropriately, resulting in improved reliability rankings through a self-learning process. These data are also used to generate valuable lists and rankings of content sources.